E/CONF.39/H/25
17 Oct.1962
ORIGINAL : ENGLISH

UNITED NATIONS CONFERENCE
ON THE APPLICATION OF SCIENCE AND TECHNOLOGY
FOR THE BENEFIT OF THE LESS DEVELOPED AREAS

Agenda Item:H.1

Food and Agriculture Organization

THE IMPACT OF ADVANCES IN STATISTICAL THEORY AND MODERN TECHNOLOGY ON AGRICULTURAL CENSUS METHODOLOGY

By C.P.G.J. SMIT,
Chief Census and Social Statistics Branch,
Statistics Division, F.A.O., Rome

Papers have been contributed to the United Nations Conference on the Application of Science and Technology for the Benefit of the Less Developed Areas as working papers for that Conference. They are published as presented by the contributors, and the contents and the views expressed are those of the contributors. (See notes overleaf.)

NOTES

1. Papers selected for inclusion in General Sessions will be available at the Conference in the following official languages of the United Nations: English, French, Russian and Spanish.

2. Papers selected for inclusion in Specialized Sessions will be available at the Conference in only one of the above official languages.

3. Summaries of all papers, as submitted by contributors, will be available at the Conference in the following official languages of the United Nations: English, French, Russian and Spanish.

4. Papers and summaries will not be available for general distribution in advance of the Conference. They will be published in advance in limited quantities only for the use of delegations of participating governments and of other authorized participants. Post-Conference publication and general distribution of the Conference Proceedings will follow United Nations practices as laid down by the United Nations Publications Board.

THE IMPACT OF ADVANCES IN STATISTICAL THEORY AND MODERN TECHNOLOGY ON AGRICULTURAL CENSUS METHODOLOGY

by

C.P.G.J. Smit, Chief Census and Social Statistics Branch, Statistics Division, FAO, Rome

1. Basic features of modern needs for statistical data

Statistical data on agricultural structure i.e. on number, distribution in space and principal characteristics of agricultural holdings, are increasingly required in less developed as well as in more advanced countries. Such information is collected in general by means of an agricultural census.

Most developing countries are at present either implementing plans for economic and social development or are in the process of formulating such plans. Agricultural censuses of a reasonable quality form one of the important pre-requisites for formulating development plans, for their implementation and for evaluation of progress in their execution. However, a typical attribute of less developed countries is paucity of statistical data. This shortage is mainly due to lack of funds, of trained personnel and of other resources.

Under these conditions it would be hardly possible to meet the increased needs if the development of statistical theory had not shown the way to collect sufficiently reliable data at the expense of only a fraction of the resources needed for carrying out a traditional agricultural census. Census taking has particularly benefitted from the advances of sampling theory and its applications.

National totals on agricultural structure and totals for large subdivisions of a country need no longer to be obtained by complete enumeration of all agricultural holdings in the country but can be estimated with a known degree of accuracy from data collected in a relatively small sample of holdings. Thus the use of sampling methods has brought the agricultural census within reach of many developing countries.

In advanced countries agricultural policies, international cooperation and the movement for economic integration require more data on agricultural structure and in greater detail than were collected before. These countries usually take their agricultural censuses by complete enumeration. Collection of all required data from all farms in the country would in many cases overburden the agricultural census questionnaires, add to the cost of the fieldwork, require a much larger number of trained enumerators and increase considerably the processing work. A growing number of countries therefore has broadened the scope of their agricultural censuses by collecting additional information from a sample of holdings only.

The increased volume of data required by many countries, at the same time, is required earlier. The introduction of electronic computers for processing agricultural censuses is bringing about substantial savings in processing time, while at the same time more complex and more comprehensive statistics can be tabulated and processing costs may be reduced.

2. Needs of developing countries

In early stages of economic growth development plans may consist only of a bundle of physical projects to be carried out during a certain period. This may be a number of irrigation wells to be dug, the quantities of improved seeds or fertilizers to be distributed or the area of land to be reclaimed during the period of the plan, together with estimates of the materials and inputs required and the estimated expenditure. Such planning takes place mainly at the national level. No too much attention is given to the mechanism of the plan as a whole. Some estimates at the national level may be sufficient in all phases of the plan. The information required at this stage can be collected by sampling and will be of rather simple character as it is limited to some basic items like land utilization, crop acreages and numbers of livestock. The data are used for making rough estimates of the total requirements as well as for a general evaluation of the effects of the plan.

A further stage of planning includes the assessment of requirements in capital, managerial and technical skills, in capital equipment and requisites and in other scarce resources which will be needed to achieve certain production goals. It includes also the establishment of priorities for scarce resources of all kinds. In using this kind of planning much more attention is focussed on the mechanism of the plan and its implementation as a whole. The targets of production and inputs are generally established for relatively large areas of the country, and the programs implemented on the same basis. The basic information in these cases should be available for the segments of the country for which distinct proposals have been made in the plan. In addition, much more detailed information is required for assessing the factors influencing production such as human, animal and mechanical power, fertilizers, irrigation facilities. If the segments are sufficiently large the data can again be supplied by a sample census.

Only at the most intensive stage of planning when targets are being set for small units, as for instance, villages, or even for individual agricultural holdings sampling cannot supply the information required and complete enumeration of all holdings in the country is necessary.

The application of sampling has furthermore enabled countries to collect information on agricultural structure under very adverse conditions. In certain parts of the world e.g. in Africa, farmers have not sufficient quantitative notions to supply statistical information by interview. In other countries, e.g. in several Near East countries, farmers are able to express quantities relating to their agricultural operations only in ambiguous units. They may be able to give the area of land in quantities of seed sown or in number of days plowing. In all these cases unambiguous information on agricultural structure can be collected only by objective measurements and counts. The cost of carrying out these objective measurements in all agricultural holdings of a country and the manpower requirements would be prohibitive even in advanced countries. The application of appropriate sampling procedures has made it possible to carry out objective measurements and counts in a limited number of holdings and thus obtain estimates for all the holdings in the country.

3. Broadening the scope of the census

The increased requirements of many countries for more and more complex data on agricultural structure could also be met on reasonable terms by the application of modern sampling methods. While the need for more data could have been satisfied by complete enumeration at the sacrifice of substantially increased costs and time, the more complicated information required in many modern agricultural censuses could be collected in most cases by application of sampling only. Several items included cannot be assessed by complete enumeration with a sufficient degree of accuracy. Examples of such items are production of crops and particularly of commodities of animal origin such as milk, eggs, meat, information on investments in agriculture.

One of the main problems in using sampling for broadening the scope of the census is the reconciliation of the estimates obtained from the sample and the results of the complete enumeration part of the census. Sample estimates are subject to sampling errors and although they may be to a lesser degree subject to non-sampling errors than the results of complete enumeration the outcome is that sample estimates of the same item are generally different from results of complete enumeration. This may lead to inconsistencies in the interrelationships between data from the two sources and may be confusing to the users.

4. Speed requirements and advance estimates

The processing of an agricultural census is a major undertaking which even under favourable circumstances will take one to two years to complete. In most cases it will take several years before the full results of the census will be available. While such delays may not be too serious for scientific studies concerned with long-term trends, for policy decisions up-to-date information is required on an increasing scale. The application of sampling to processing has greatly helped to meet these requirements. Advance results of sufficient accuracy can be prepared by tabulation of the returns from a relatively small sample of agricultural holdings.

The basic operations of processing a census are simple, and therefore often neglected or underrated. In view of large amount of data to be processed in censuses mechanical means have been introduced to speed up the processing. Many statistical offices are today equipped with punch-card machines which make it possible to cope with the large amount of census data. The most powerful tools for speeding up the processing operations of censuses are electronic computers. Computers are used already by a few national statistical offices, mainly in the more developed countries, because they found it impossible to cope with the increasing amount of processing work. Electronic computers can substitute a large part of the conventional punch-card machines. They are speeding up the mechanical part of the processing considerably.

There is one very important although less publicized possibility offered by computers. Formerly the questionnaires had to undergo a time consuming process of manual checking and editing before the preparation of punched cards. Now all these checks including those on arithmetical and other relations can be made by the computer. The computer indicates the wrong or improbable

relations which then have to be corrected manually. It is obvious that this method of mechanical editing by computers will not only increase the speed of processing but also the accuracy of the tabulations.

Furthermore, the computer is also able to correct minor errors automatically, if these corrections can be made according to some rules known from experience. The manual corrections can be reduced in this way to errors to be corrected by reference to the questionnaires.

The present status of the technical development of processing of statistical data has not yet decreased the time needed for the preparation of the input data. Punching and verifying of cards or of paper tapes still have to be prepared in the conventional way. The preparation of the data for the computers, in general the punching and verifying of cards, remains the main bottleneck for the processing of census data.

Several types of machines have been developed for automatically reading documents. Still further experimentation is needed before these machines can be used for the processing of census data in all parts of the world. The most interesting reading machine for statistical work has been developed and used by the US Bureau of the Census (FOSDIC, Film Optical Sensing Device for Input to Computers).

5. Broadening tabulation programs

Sample tabulation is being used not only for the preparation of advance estimates but also for other purposes. First all additional tabulations may be needed which would be too costly or too time consuming to prepare on the basis of complete tabulation. If such additional data are not required for small administrative units sample tabulation may be the answer to the problem.

In some recent censuses sampling has been used also for the preparation of cross tabulations. The necessary caution has to be exercised in this case because the number of units in each cell of extensive cross-tabulation may become too small for obtaining sufficiently reliable results which may be subject to high sampling errors.

Sample tabulation is also used as a last resource for obtaining results from complete enumeration in cases where resources, staff or equipment have proved to be inadequate for carrying out the full tabulation program. Processing requirements tend to be underestimated in planning the census and this case may be more frequent than generally is assumed.

The use of electronic computers has opened further possibilities for broadening the tabulation program. Because the speed of electronic computers is considerably higher than that of other processing tools, it is obvious that during the same period of time more tables can be prepared by computers than by other machines. Additional tables may be prepared for additional subgroups or additional territorial subunits. It is especially convenient to prepare cross-tabulations with computers. All items of one unit, for example, an agricultural holding, can be stored in the computer and can be related to each other.

Moreover, while preparing the usual census tables the computer, without using practically any additional time, may calculate derived figures like means, percentages, standard deviations, index numbers, etc.

6. Improved accuracy

Modern sampling techniques have also contributed to increasing the accuracy of agricultural census data. This is being done at various stages of the census work. Firstly questionnaires and enumeration procedures are being tested in a sample of holdings. This will improve the formulation of the questionnaire and indicate the best procedures to be used in the actual census. The development of sampling theory and practice has also permitted the building of checking devices into the census. The data from all agricultural holdings collected by the interview method are then checked by objective measurements and counts in a sample of holdings. If the census is already carried out in a large sample the checking may take place in a sub-sample.

The checking operations are carried by the enumerators of the complete census or by special investigators. In the latter case investigators can be used who are more skilled and better trained than the general enumerators. So far the use of quality checks built into the agricultural census itself has found limited applications. One of the reasons may be that during a complete enumeration all energy of the census takers has to be concentrated on the main job in hand and that it is difficult to give at the same time adequate attention to the quality checks. Preference has been given to checking the census returns by post-enumeration checks, undertaken when the main enumeration work was over. The purpose of these checks is twofold. First of all they give an indication of the degree of accuracy of the data collected in the main enumeration. This information naturally is of particular importance for all users of the data. The post enumeration checks also supply indication of the shortcomings of the enumeration methods, which information in its turn can be used for improvement of further censuses and surveys.

The introduction of sampling methods furthermore has led to substantial savings on the cost of controlling the quality of processing work of censuses at its various stages. It must be recognized however that in agricultural censuses quality control of processing is much more complicated than in population censuses. The latter deals with units (individuals only) while agricultural censuses largely are concerned with quantities. Nevertheless, sampling can be applied also in these censuses successfully for various purposes as for instance detecting at an early stage of the processing operators whose performance is inadequate. It is however clear that the principles of industrial quality control will have to be substantially modified before they can be used effectively in the control of the quality of processing agricultural censuses.

7. Rationalisation

Testing of census questionnaires and enumeration procedures by sampling methods has contributed not only to improving the accuracy of census results but also has been conducive to a more rational census planning.

Agricultural censuses in most countries are being taken at long intervals. Experience from one census may have been partly lost or may have become obsolete by the time the next census is due. Countries taking the census for the first time could not rely at all on national experience. In all these cases a certain amount of experimentation needs to be carried out before countries embark on the major census operation. In the past census plans were mostly based on guess work. Any pre-testing was done on purposively selected samples, and the information thus obtained was mostly of a qualitative type or gave only rough quantitative indications.

The introduction of random sampling methods in pre-testing census questionnaires and procedures has provided census designers with a tool for making close estimate of the various quantitative data required for a rational planning of the census operations.

While the development of statistical theory has had a considerable impact on the field work of agricultural censuses, the development of electronic computers is on the way to revolutionising the processing of census data. The computers offer not only high speed for sorting, counting, adding, etc., but also new storage techniques for statistical data by using magnetic tapes, discs, drums, etc. Computers are able to perform quite different operations without manual interruption. These operations had to be performed previously by different types of punch-card machines which required frequent manual interventions. Because of this the use of electronic computers in processing census data is introducing a completely new concept of census processing.

Those planning, directing and executing the census need therefore to be acquainted with the principal working techniques of electronic computers. For instance, questionnaire design and punch-card design have to be adapted to computer methods, otherwise computer time will be wasted. It is also important that the officers in charge of the computer are associated with planning and organization of the census from the earliest stages. In a way, electronic computers represent the centre of an automatic factory and require thus that all events are programmed. Improvisation which plays an important role in manual processing and to a certain extent is still practised when using punch-card machines, cannot be permitted at all in using electronic computers. Computer processing of agricultural census-returns thus tends to rationalize the whole process of census taking.

8. Logistic problems arising from technological development

When in the last five years electronic computers came on the market, one part of the statistical offices, mainly in the more developed countries, were already equipped with punch-card machines, whereas the other part, mainly in the less developed countries, had no punch-card equipment available. Because the computers usable for statistical purposes are using punched cards as input media, it was natural that at first only those offices planned to install and installed computers which had already gained a large experience in the handling of punch-cards. Nevertheless, even these offices needed a careful preparation for the setting up of the computers, in most cases requiring several years. The preparation includes construction or reconstruction of a building for the installation, training of programmers and operators, reorganization of the work flow, preparation of machine programs, checking the machine programs on other computer installations.

At present computers are available under such terms that it is economically feasible for many statistical offices having no punch-card installation to instal directly an electronic computer without going first through the full experience of using a punch-card installation. So far as is known no statistical office has gone this way as yet. The main problem seems to be that the difficulties to be encountered in installing a computer in statistical offices with no previous experience in punch-card handling are unknown. Apparently there would be a large scope for international assistance in this field. Past international assistance in data processing by UN and FAO has been concentrated on manual, semi-manual and punch-card methods.

A handbook on Data Processing Methods has been published jointly by the two Organizations and a training centre on processing of agricultural census data has been held for the Near East countries which were particularly in need of such assistance. Only a pilot project for transferred processing in which the Libyan agricultural census data were processed by an electronic computer in Germany was concerned with the processing techniques by computers. This project was organized by FAO, because more than 20 countries from all over the world have shown interest in having their census data processed by computers in another country than that where the data had been collected. Also in this case no experience was available for dealing with the technical and logistic problems involved in such an undertaking.

The pilot project was carried out to fill in these gaps in knowledge. FAO has explored the facilities for the use of electronic computers for such a pilot project. The United Kingdom of Libya offered to provide the data of their 1960 agricultural census for processing under FAO supervision. The Provisional International Computation Centre, located in Rome, made arrangements with the "Institut für Praktische Mathematik" at Darmstadt for the necessary computer time and the technical services of the computer staff. The Federal Office of Statistics, Wiesbaden, Federal Republic of Germany, undertook the punching and verification of Libyan data and gave other assistance. FAO helped to plan and organize the work and provided the necessary supervision and control of the project. As result of this cooperation the Government of Libya received the complete final results of its agricultural census, covering 150 000 agricultural holdings early in 1962. This is a remarkable achievement for a country which never took an agricultural census before.

The significance of the project surpassed its national importance by the technical and logistic data it yielded for application in other similar cases. The main findings have been published in a technical report by the International Computation Centre, Rome. 1/

1/ Dilwali, C.K. Report on the Pilot Project for Central Tabulation of the Agricultural Census Data for the United Kingdom of Libya by Electronic Computers, Rome (1962).

The project has given hope that the development in electronic computation may be used to a much greater advantage in ensuring speedy tabulation of the 1970 censuses of agriculture in several less developed countries. International assistance in this field might well take the form of organizing regional processing centres provided with electronic computers. It is true that many developed countries are still experiencing difficulties in tabulating their own census data using such computers, this is due to a number of reasons. The number of items included in their questionnaires is very large and they have a heavy program of tabulation and cross-tabulations to carry out. By comparison the number of items and the program of tabulation in the censuses of most developing countries is very modest. Secondly, although the use of electronic computers for processing mass statistical data is still at an exploratory stage, big advances are expected to take place by the time the 1970 census is taken. Thirdly, even if there are no savings in cost there undoubtedly will be very substantial savings in time. These reasons lead to a belief that modern technology can be effectively used in speedy tabulation and publication of the 1970 censuses of agriculture.